

Klasifikasi Sentimen *Google Play Store* Aplikasi ChatGPT Berbahasa Indonesia Berbasis IndoBERT

Ichsani Mursidah¹, Remi Sanjaya², Bambang Yulianto³, Dhian Sweetania⁴, Puji Sularsih⁵
^{1,4,5}Prodi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma, Jakarta, Indonesia

²Prodi Informatika, Fakultas Teknologi Industri, Universitas Gunadarma, Jakarta, Indonesia

³Prodi Industri, Fakultas Teknologi Industri, Universitas Gunadarma, Jakarta, Indonesia

¹Mursidah@staff.gunadarma.ac.id, ²remi@staff.gunadarma.ac.id,

³bambang_yulianto@staff.gunadarma.ac.id, ⁴dhian_sweetania@staff.gunadarma.ac.id,

⁵puji@staff.gunadarma.ac.id

Submit : 30 Nov 2025 | Diterima : 25 Des 2025 | Terbit : 28 Des 2025

ABSTRAK

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi ChatGPT berbahasa Indonesia dengan memanfaatkan metode *IndoBERT*. Sentimen pengguna diklasifikasikan ke dalam tiga kelas, yaitu positif, negatif, dan netral. Data yang digunakan dalam penelitian ini berjumlah 25.111 ulasan yang diperoleh dari *Google Play Store*. Dataset tersebut kemudian melalui tahapan *pra-pemrosesan* teks yang meliputi *text cleaning*, *tokenization*, penghapusan *stopword*, normalisasi, serta *stemming*. Metode *IndoBERT* diterapkan pada proses pelabelan sekaligus klasifikasi sentimen, sementara kinerja model dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa model *IndoBERT* mampu mencapai nilai akurasi sebesar 89%, presisi 87%, *recall* 89%, dan *F1-score* sebesar 88%. Temuan ini mengindikasikan bahwa *IndoBERT* memiliki performa yang baik dan efektif dalam melakukan klasifikasi sentimen pada teks berbahasa Indonesia. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan kajian analisis sentimen berbasis Bahasa Indonesia serta menjadi referensi dalam upaya peningkatan kualitas dan pengalaman pengguna pada aplikasi ChatGPT maupun aplikasi serupa.

Kata Kunci: Analisis Sentimen, *Indobert*, *Deep Learning*, Ulasan Pengguna, Bahasa Indonesia

PENDAHULUAN

Perkembangan kecerdasan buatan (*Artificial Intelligence/AI*) telah membawa pengaruh yang signifikan terhadap berbagai aspek kehidupan, khususnya dalam bidang komunikasi dan penyediaan informasi. Salah satu penerapan AI yang saat ini banyak digunakan oleh masyarakat adalah ChatGPT, yaitu model bahasa berbasis *transformer* yang dikembangkan oleh OpenAI. Aplikasi ini memungkinkan terjadinya interaksi berbasis bahasa alami (*natural language*) secara mendalam, mulai dari pencarian informasi, penulisan teks, hingga sarana hiburan, sehingga memberikan kemudahan bagi pengguna dalam aktivitas sehari-hari.

Seiring dengan meningkatnya penggunaan ChatGPT, persepsi dan tanggapan pengguna terhadap layanan tersebut menjadi aspek penting untuk dikaji. Pemahaman terhadap opini pengguna secara objektif memerlukan pendekatan yang sistematis, salah satunya melalui analisis sentimen. Analisis sentimen merupakan teknik yang digunakan untuk mengklasifikasikan opini atau pendapat ke dalam kategori positif, negatif, atau netral berdasarkan data tekstual yang dihasilkan oleh pengguna.

Dalam konteks analisis sentimen berbahasa Indonesia, salah satu model yang terbukti memiliki kinerja yang baik adalah *IndoBERT* (*Bidirectional Encoder Representations from Transformers*). Model ini merupakan pengembangan *BERT* yang telah dilatih secara khusus

menggunakan korpus Bahasa Indonesia, sehingga mampu menangkap konteks linguistik lokal dengan lebih baik. Jika dibandingkan dengan model multibahasa seperti *mBERT*, *IndoBERT* menunjukkan keunggulan dalam memahami struktur bahasa dan karakteristik semantik Bahasa Indonesia.

Penelitian yang dilakukan oleh Nurhasiyah *et al.* (2025) menunjukkan bahwa *IndoBERT* mampu mencapai tingkat akurasi sebesar 84% dengan nilai *precision* 75%, *recall* 80%, dan *F1-score* 78% dalam tugas klasifikasi sentimen. Berdasarkan capaian tersebut, *IndoBERT* dipilih sebagai metode utama dalam penelitian ini untuk menganalisis sentimen ulasan pengguna ChatGPT berbahasa Indonesia secara lebih akurat dan representatif.

TINJAUAN PUSTAKA

Analisis Sentimen

Analisis sentimen, yang juga dikenal sebagai *opinion mining*, merupakan kombinasi antara *text mining* dan *data mining* yang digunakan untuk mengidentifikasi serta mengevaluasi opini, sentimen, dan emosi yang terkandung dalam suatu teks. Analisis sentimen berfokus pada pengkajian pendapat, sikap, evaluasi, serta emosi yang disampaikan oleh penulis atau pembicara terhadap topik tertentu, seperti produk, layanan, organisasi, individu, maupun aktivitas lainnya (Mailo & Lazuardi, 2019).

Google Play Store

Google Play Store merupakan platform distribusi digital yang dikembangkan dan dikelola oleh Google untuk perangkat berbasis sistem operasi Android. Selain sebagai media distribusi aplikasi, *Google Play Store* juga menyediakan fitur pemberian ulasan dan penilaian (*rating*) terhadap aplikasi yang telah digunakan. Fitur tersebut memungkinkan pengguna untuk menyampaikan pengalaman serta opini mereka, sehingga dapat menjadi referensi bagi pengguna lain sebelum mengunduh aplikasi tertentu (Herlinawati *et al.*, 2020).

ChatGPT

ChatGPT merupakan model bahasa generatif yang dikembangkan menggunakan arsitektur *Transformer* dan dirancang untuk berbagai tugas *Natural Language Processing (NLP)* (Marquis *et al.*, 2024). ChatGPT dilatih menggunakan kumpulan data berskala besar yang bersumber dari internet, seperti halaman web, artikel, dan konten media sosial dalam berbagai bahasa, sehingga memiliki cakupan pengetahuan yang luas (Hassani & Silva, 2023). Menurut Kalla dan Kuraku (2023b), ChatGPT diimplementasikan melalui tiga model utama, yaitu *Supervised Fine-Tuning (SFT)*, *Reward Model (RM)*, serta *SFT* berbasis *Proximal Policy Optimization (PPO)*.

Google Colaboratory

Google Colaboratory atau *Google Colab* merupakan platform berbasis *cloud* yang disediakan oleh Google untuk menulis dan menjalankan kode Python secara daring. Platform ini memungkinkan pengguna untuk menjalankan program tanpa memerlukan perangkat keras dengan spesifikasi tinggi. Keunggulan *Google Colab* antara lain bersifat gratis, mudah diakses, mendukung komputasi berbasis *cloud*, serta menyediakan dukungan *GPU (Graphics Processing Unit)* dan *TPU (Tensor Processing Unit)*. Selain itu, *Google Colab* terintegrasi dengan *Google Drive* dan mendukung kolaborasi antar pengguna. Platform ini juga memungkinkan eksekusi *Jupyter Notebook* secara langsung melalui peramban web tanpa memerlukan konfigurasi tambahan (Febrywinata, 2024).

Artificial Intelligence

Artificial Intelligence (AI) atau kecerdasan buatan merupakan cabang ilmu komputer yang berfokus pada pengembangan sistem atau mesin yang mampu meniru kemampuan berpikir dan belajar manusia. *AI* mencakup berbagai metode dan teknologi, seperti *machine learning*, *deep learning*, *natural language processing*, *computer vision*, dan *robotics*. Melalui penerapan *AI*, sistem komputer dapat mempelajari pola berpikir manusia, menyimpan pengetahuan yang diperoleh, serta mengaplikasikannya untuk menyelesaikan tugas-tugas tertentu secara cerdas (Sitanggang, Rianto, & Sibagariang, 2019).

Machine Learning

Machine learning merupakan salah satu bidang dalam kecerdasan buatan yang memungkinkan mesin komputer mempelajari data untuk menghasilkan prediksi atau solusi tanpa pemrograman eksplisit. Penerapan *machine learning* sangat bergantung pada ketersediaan data, yang umumnya dibagi menjadi data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun dan melatih model, sedangkan data uji berfungsi untuk mengukur kinerja model dalam menghadapi data baru yang belum pernah dipelajari sebelumnya (Fikriya, Irawan, & Soetrisno, 2017).

Deep Learning

Deep learning merupakan bagian dari *machine learning* yang berbasis pada *Artificial Neural Network (ANN)* atau jaringan saraf tiruan. Metode ini terdiri atas beberapa lapisan (*layers*), yaitu *input layer*, *hidden layer*, dan *output layer*, yang digunakan untuk mempelajari representasi data secara hierarkis. *Deep learning* dirancang untuk meniru cara kerja otak manusia dalam menganalisis data dan mengambil keputusan. Untuk meningkatkan performanya, metode ini memanfaatkan algoritma *ANN* yang terinspirasi dari jaringan biologis otak manusia (Peryanto *et al.*, 2019).

Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers (BERT) merupakan model *pre-trained Natural Language Processing* yang dikembangkan oleh Google untuk meningkatkan pemahaman konteks bahasa dalam berbagai aplikasi, termasuk mesin pencari. Model *BERT* bersifat *open source* dan dapat dikembangkan menggunakan teknik *deep learning*. Proses pelatihan *BERT* terdiri atas dua tahap, yaitu *pre-training* dan *fine-tuning*. Pada tahap *pre-training*, *BERT* dilatih menggunakan dua tugas *unsupervised*, yaitu *Masked Language Modeling (MLM)* dan *Next Sentence Prediction (NSP)*.

IndoBERT

IndoBERT merupakan model *pre-trained* berbasis *BERT* yang dikembangkan khusus untuk Bahasa Indonesia. Model ini dilatih menggunakan korpus *Indo4B* yang terdiri atas sekitar 4 miliar kata dan 250 juta kalimat berbahasa Indonesia, yang bersumber dari berbagai media seperti Tempo, Kompas, dan Wikipedia. Sama seperti *BERT*, *IndoBERT* menggunakan arsitektur *multi-layer bidirectional transformer*. *IndoBERT* memiliki beberapa varian model, antara lain *IndoBERT-LiteBase*, *IndoBERT-Base*, dan *IndoBERT-Large*, yang masing-masing memiliki perbedaan jumlah parameter, lapisan, dan ukuran *hidden layer* (Wilie *et al.*, 2020).

Confusion Matrix

Confusion matrix merupakan metode evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan kelas sebenarnya (*actual class*) dan kelas hasil prediksi (*predicted class*). *Confusion matrix* menjadi dasar dalam perhitungan berbagai metrik evaluasi, seperti *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion matrix* terdiri atas empat komponen utama, yaitu *true positive (TP)*, *true negative (TN)*, *false positive (FP)*, dan *false negative (FN)*. Nilai-nilai tersebut memberikan informasi yang lebih rinci mengenai kemampuan model dalam mengklasifikasikan data, termasuk tingkat kesalahan yang dihasilkan. Oleh karena itu, *confusion matrix* banyak digunakan sebagai alat evaluasi dalam penelitian klasifikasi, khususnya pada analisis sentimen berbasis teks (Harun & Ananda, 2021).

1. Accuracy

Nilai *accuracy* diperoleh dari perbandingan jumlah prediksi yang benar terhadap total data yang diuji. Metrik ini memberikan gambaran umum mengenai performa model, namun kurang representatif apabila dataset tidak seimbang.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2. Precision

Metrik ini menunjukkan seberapa besar proporsi data yang diprediksi positif benar-benar termasuk ke dalam kelas positif.

$$Precision = \frac{TP}{TP + FP}$$

3. *Recall*

Nilai *recall* menunjukkan seberapa besar proporsi data positif yang berhasil teridentifikasi dengan benar oleh model.

$$Recall = \frac{TP}{TP + FN}$$

4. *F1-Score*

F1-score merupakan metrik evaluasi yang menggabungkan nilai *precision* dan *recall* dalam bentuk rata-rata harmonik.

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall}$$

METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan penelitian yang disusun secara sistematis untuk menganalisis sentimen ulasan pengguna ChatGPT berbahasa Indonesia. Secara umum, alur penelitian terdiri atas beberapa tahapan utama, mulai dari pengumpulan data hingga evaluasi kinerja model.

Pengumpulan Data (*Data Scraping*)

Data ulasan pengguna diperoleh melalui teknik *web scraping* menggunakan bahasa pemrograman Python. Proses pengambilan data dilakukan pada platform *Google Play Store* sebagai sumber utama ulasan aplikasi ChatGPT. Data yang dikumpulkan berupa teks ulasan pengguna dalam Bahasa Indonesia yang merepresentasikan opini dan pengalaman penggunaan aplikasi.

Pra-pemrosesan Data (*Pre-processing*)

Tahapan *pra-pemrosesan* dilakukan untuk membersihkan dan menyiapkan data teks agar dapat diproses secara optimal oleh model. Proses ini dilakukan secara berurutan dengan tahapan sebagai berikut:

1. Case Folding, yaitu mengubah seluruh teks menjadi huruf kecil untuk menghindari perbedaan makna akibat variasi penulisan huruf.
2. Cleansing, yaitu menghapus karakter yang tidak relevan seperti tanda baca, angka, simbol, dan *emoji*.
3. Tokenization, yaitu memecah teks menjadi unit kata atau token.
4. Stopword Removal, yaitu menghilangkan kata-kata umum yang tidak memiliki makna signifikan terhadap sentimen.
5. Normalization, yaitu menyamakan bentuk kata tidak baku menjadi bentuk baku.
6. Stemming, yaitu mengubah kata ke bentuk dasar untuk mengurangi variasi kata yang memiliki makna serupa.

Pelabelan Data (*Data Labelling*)

Setelah melalui proses *pra-pemrosesan*, data ulasan diberi label sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Proses pelabelan dilakukan secara manual atau semi-otomatis untuk memastikan akurasi dan konsistensi label, sehingga kualitas dataset tetap terjaga.

Pembagian Data (*Data Splitting*)

Data yang telah diberi label selanjutnya dibagi menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*) dengan proporsi tertentu. Pembagian ini bertujuan agar model dapat dilatih secara optimal sekaligus diuji kemampuannya dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.

Penerapan Model *IndoBERT*

Penerapan model *IndoBERT* dilakukan melalui beberapa tahapan sebagai berikut:

1. **Inisialisasi Model**, yaitu mengimpor dan mengonfigurasi model *IndoBERT* yang tersedia pada pustaka *Hugging Face*. Model *pre-trained* ini digunakan sebagai dasar analisis sentimen.

2. **Fine-tuning**, yaitu proses pelatihan ulang model *IndoBERT* menggunakan dataset ulasan pengguna yang telah diproses, sehingga model mampu menyesuaikan diri dengan konteks dan karakteristik bahasa pada data penelitian.
3. **Inferensi**, yaitu proses klasifikasi sentimen terhadap setiap ulasan pengguna menggunakan model yang telah dilatih.
4. **Post-processing**, yaitu pengolahan hasil klasifikasi agar tersusun secara sistematis dan siap digunakan untuk analisis lanjutan.

Proses Pelatihan (*Training*)

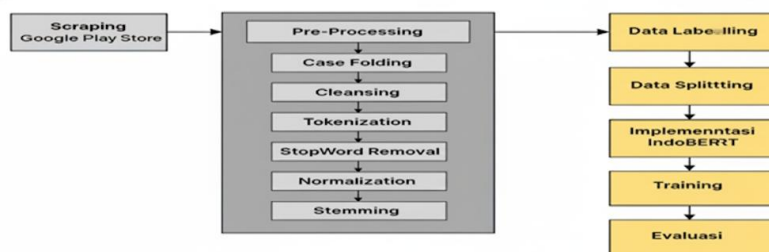
Pada tahap *training*, model *IndoBERT* dilatih menggunakan data latih selama beberapa *epoch* untuk mengoptimalkan parameter model. Selama proses pelatihan, performa model dipantau melalui visualisasi seperti grafik *loss* dan akurasi pada data latih dan data uji guna memastikan model tidak mengalami *overfitting* maupun *underfitting*.

Evaluasi Performa Model

Evaluasi kinerja model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil evaluasi tersebut digunakan untuk menilai tingkat keefektifan model *IndoBERT* dalam melakukan klasifikasi sentimen ulasan pengguna berbahasa Indonesia, serta dibandingkan dengan hasil penelitian sebelumnya sebagai bahan analisis dan pembahasan.

HASIL DAN PEMBAHASAN

Secara keseluruhan, tahapan-tahapan ini membentuk kerangka kerja penelitian yang sistematis dan ditampilkan dalam Gambar 1 sebagai alur proses.



Gambar 1. Tahap Penelitian

Pengumpulan Data (*Data Scraping*)

Pada tahap ini, penelitian difokuskan pada proses pengumpulan data berupa ulasan pengguna aplikasi ChatGPT yang tersedia pada platform *Google Play Store*. Pengambilan data dilakukan menggunakan teknik *web scraping* dengan memanfaatkan pustaka *google-play-scraper*. Pustaka tersebut memungkinkan proses pengambilan ulasan dilakukan secara otomatis berdasarkan parameter tertentu, seperti jumlah data dan iterasi pengambilan. Penelitian ini menargetkan pengumpulan sekitar 25.000 ulasan pengguna, dengan jumlah pengambilan sebanyak 200 ulasan pada setiap iterasi. Proses ini dilakukan secara bertahap hingga jumlah ulasan yang ditentukan dapat terpenuhi. Hasil dari tahap *scraping* menghasilkan kumpulan data teks ulasan berbahasa Indonesia yang siap digunakan pada tahapan analisis sentimen selanjutnya.



	content	score
0	cukup bagus	0
1	uang saya hilang tolong kembalikan chat gpt	1
2	👍keren	5
3	mantap	5
4	cukup bagus untuk buatan manusia. tapi kalo buatan tuhan pasti itu sempurna 🙏	4
...		
25106	halo lin chatgpt, kenapa sekarang chatgpt ku jadi "seperti nya ada kesadaran", tapi pas di pencet ulang lagi masih sama "seperti nya ada kesadaran" lolong di lindak lanjut yai	2
25107	grate 🙏🙏	7
25108	yg bikin masalah yll	5
25109	knp chat gpt gue gabisa akses motor torus anjar gue kira hp gue tapi nyata agh nyie knp wab yang belang tlc. gaada bukti gue tangsung kerm omali ku jmlahnya	2
25110	suka kurang jawababnya (terpotong) ga kek ai yang lain	1

Gambar 2. Hasil *Data Scraping*

Data ulasan yang diperoleh merepresentasikan persepsi dan pengalaman langsung pengguna terhadap aplikasi ChatGPT. Oleh karena itu, dataset ini menjadi dasar yang kuat dalam menganalisis tingkat kepuasan pengguna sekaligus mengevaluasi kinerja model klasifikasi sentimen yang dibangun. Keberagaman opini dalam ulasan tersebut juga memberikan gambaran yang komprehensif mengenai respons pengguna terhadap fitur dan performa aplikasi ChatGPT.

Pra-pemrosesan Data (*Pre-processing*)

Tahapan ini dilakukan karena data yang diperoleh masih berada dalam kondisi mentah, tidak terstruktur, serta mengandung berbagai elemen yang dapat mengganggu proses analisis sentimen.

1. *Case Folding*

Tahap awal dalam proses *pra-pemrosesan* teks adalah *case folding*, yaitu proses mengonversi seluruh huruf dalam teks menjadi huruf kecil (*lowercase*). Langkah ini bertujuan untuk menghilangkan perbedaan antara huruf kapital dan huruf kecil yang tidak memiliki makna semantik dalam analisis sentimen. Dengan penerapan *case folding*, variasi penulisan seperti “Surya”, “surya”, dan “SURYA” akan diperlakukan sebagai kata yang sama. Selain itu, proses ini membantu mengurangi kompleksitas kosakata dalam data teks karena seluruh bentuk penulisan telah diseragamkan.

	Teks Asli	a. Case Folding
0	cukup bagus	cukup bagus
1	uang Saya hilang tolong kembalikan chat gpt	uang saya hilang tolong kembalikan chat gpt
2	👍 keren	👍 keren
3	mantap	mantap
4	cukup bagus untuk buatan manusia ,tapi kalo buatan tuhan pasti itu sempurna 😊	cukup bagus untuk buatan manusia ,tapi kalo buatan tuhan pasti itu sempurna 😊

Gambar 3. Hasil *Case Folding*

2. *Cleansing*

Tahapan berikutnya adalah *cleansing*, yaitu proses pembersihan teks dari berbagai elemen yang tidak relevan atau bersifat *noise*, seperti *Uniform Resource Locator (URL)*, tanda baca, karakter khusus, serta angka. Proses ini menghasilkan teks ulasan yang lebih bersih dan terfokus pada informasi yang memiliki nilai analisis.

a. Case Folding	b. Cleansing
cukup bagus	cukup bagus
uang saya hilang tolong kembalikan chat gpt	uang saya hilang tolong kembalikan chat gpt
👍 keren	keren
mantap	mantap
cukup bagus untuk buatan manusia ,tapi kalo buatan tuhan pasti itu sempurna 😊	cukup bagus untuk buatan manusia tapi kalo buatan tuhan pasti itu sempurna

Gambar 4. Hasil *Cleansing*

Dengan menghilangkan elemen-elemen tersebut, data yang digunakan pada tahap selanjutnya menjadi lebih terstruktur dan tidak terganggu oleh karakter yang tidak memberikan kontribusi terhadap penentuan sentimen.

3. *Tokenization*

Setelah melalui proses *cleansing*, tahapan selanjutnya adalah *tokenization*, yaitu proses memecah teks menjadi unit-unit kecil yang disebut token, umumnya berupa kata per kata. Pada tahap ini, setiap kalimat diubah menjadi daftar kata yang merepresentasikan komponen penyusun teks. Hasil *tokenization* menjadi dasar bagi tahapan *pra-pemrosesan* lanjutan, seperti *stopword removal* dan *stemming*, karena setiap token dianggap sebagai satuan informasi yang akan dianalisis dan dipelajari oleh model klasifikasi sentimen.

b. Cleansing	c. Tokenization
cukup bagus	[cukup, bagus]
uang saya hilang tolong kembalikan chat gpt	[uang, saya, hilang, tolong, kembalikan, chat, gpt]
keren	[keren]
mantap	[mantap]
cukup bagus untuk buatan manusia tapi kalo buatan tuhan pasti itu sempurna	[cukup, bagus, untuk, buatan, manusia, tapi, kalo, buatan, tuhan, pasti, itu, sempurna]

Gambar 5. Hasil *Tokenization*

4. *Stopword Removal*

Tahap *stopword removal* merupakan proses penghapusan kata-kata umum yang sering muncul dalam teks, seperti “dan”, “di”, “yang”, dan “adalah”, namun tidak memiliki kontribusi signifikan terhadap makna sentimen. Keberadaan *stopword* dapat menambah *noise* dalam data sehingga perlu dihilangkan untuk meningkatkan fokus analisis pada kata-kata yang bermakna.

Dalam penelitian ini, proses *stopword removal* dilakukan menggunakan pustaka *Sastrawi* yang menyediakan daftar *stopword* khusus Bahasa Indonesia.

c. Tokenization	d. Stopword Removal
[cukup, bagus]	[cukup, bagus]
[uang, saya, hilang, tolong, kembalikan, chat, gpt]	[uang, hilang, kembalikan, chat, gpt]
[keren]	[keren]
[mantap]	[mantap]
[cukup, bagus, untuk, buatan, manusia, tapi, kalo, buatan, tuhan, pasti, itu, sempurna]	[cukup, bagus, buatan, manusia, kalo, buatan, tuhan, sempurna]

Gambar 6. Hasil *Stopword Removal*

5. Normalization

Normalisasi bertujuan untuk mengubah kata-kata tidak baku atau bentuk bahasa informal yang umum digunakan dalam percakapan sehari-hari dan media sosial menjadi bentuk baku sesuai dengan Kamus Besar Bahasa Indonesia (*KBBI*). Melalui normalisasi, variasi penulisan seperti “gk”, “ga”, atau “tdk” diseragamkan menjadi “tidak”, sehingga model dapat memahami makna kata secara lebih konsisten.

d. Stopword Removal	e. Normalization
[cukup, bagus]	[cukup, bagus]
[uang, hilang, kembalikan, chat, gpt]	[uang, hilang, kembalikan, chat, gpt]
[keren]	[keren]
[mantap]	[mantap]
[cukup, bagus, buatan, manusia, kalo, buatan, tuhan, sempurna]	[cukup, bagus, buatan, manusia, kalo, buatan, tuhan, sempurna]

Gambar 7. Hasil *Normalization*

6. Stemming

Tahap *stemming* bertujuan untuk mereduksi kata ke bentuk dasarnya dengan menghilangkan imbuhan atau variasi morfologis. Proses ini memastikan bahwa kata-kata seperti “berlari”, “pelari”, dan “berlari-lari” dianalisis sebagai satu entitas yang sama, yaitu “lari”. Dengan menerapkan *stemming*, kata-kata turunan seperti “membantu”, “dibantu”, dan “bantuan” dipetakan ke bentuk dasar “bantu”. Tahapan ini merupakan langkah terakhir dalam proses *pra-pemrosesan* sebelum data digunakan pada tahap pelabelan dan pelatihan model analisis sentimen.

e. Normalization	f. Stemming
[cukup, bagus]	[cukup, bagus]
[uang, hilang, kembalikan, chat, gpt]	[uang, hilang, kembali, chat, gpt]
[keren]	[keren]
[mantap]	[mantap]
[cukup, bagus, buatan, manusia, kalo, buatan, tuhan, sempurna]	[cukup, bagus, buat, manusia, kalo, buat, tuhan, sempurna]

Gambar 8. Hasil *Stemming*

Data Labelling

Setelah data ulasan melalui tahap *pra-pemrosesan*, dilakukan proses pelabelan sentimen berdasarkan skor *rating* yang diberikan pengguna. Setiap ulasan dikategorikan ke dalam tiga kelas, yaitu positif, netral, dan negatif, sehingga data siap digunakan dalam skema *supervised learning*. Hasil pelabelan menunjukkan distribusi sentimen terdiri atas 20.545 ulasan positif, 3.391 ulasan negatif, dan 1.175 ulasan netral.

```
Distribusi Label:
label
positif    20545
negatif    3391
netral     1175
Name: count, dtype: int64
```

Gambar 9. Hasil *Data Labelling*

Data Splitting

Data yang telah diberi label selanjutnya dibagi menjadi data latih (*training set*) dan data uji (*test set*) menggunakan fungsi *train_test_split* dari pustaka *scikit-learn*. Pembagian dilakukan dengan rasio 80:20, sehingga dari total 25.111 data diperoleh 20.088 data latih dan 5.023 data uji.

Penggunaan parameter *stratify* memastikan distribusi kelas pada kedua subset tetap proporsional dan representatif.

Tabel 1. Hasil Data Splitting

Jenis Data	Positif	Negatif	Netral	Total
Data Latih (<i>Training Set</i>)	16.435	2.713	940	20.088
Data Uji (<i>Testing Set</i>)	4.110	678	235	5.023
Total Keseluruhan	20.545	3.391	1.175	25.111

Penerapan Model *IndoBERT*

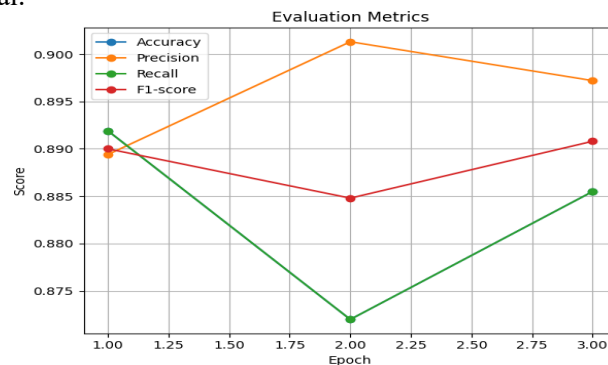
Tahap ini merupakan langkah awal dalam membangun model klasifikasi sentimen berbasis *transformer* dengan memanfaatkan model *pre-trained IndoBERT*. Seluruh data ulasan telah dikonversi ke dalam bentuk vektor dan label numerik, sehingga siap digunakan pada proses pelatihan model untuk klasifikasi sentimen Bahasa Indonesia.



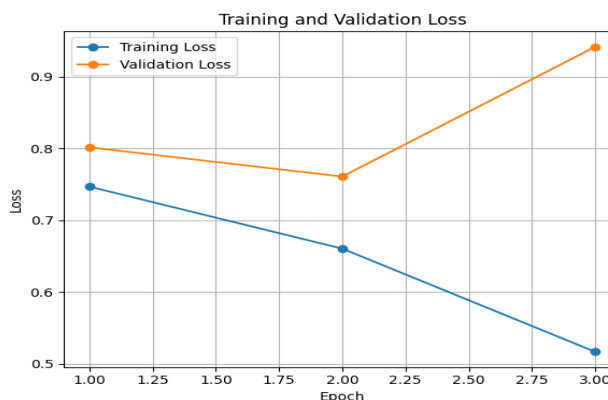
Gambar 10. Hasil Proses Load Model

Training

Dataset ulasan pengguna yang telah diberi label dikonversi ke dalam bentuk numerik, dengan skema negatif sebagai 0, netral sebagai 1, dan positif sebagai 2. Data diproses menggunakan format *Dataset* dan dilatih dengan kelas *Trainer* menggunakan parameter yang telah ditentukan. Selama pelatihan, dilakukan pemantauan terhadap *training loss*, *validation loss*, akurasi, *precision*, *recall*, dan *F1-score* pada setiap *epoch*. Berdasarkan gambar 11, hasil pelatihan menunjukkan penurunan *training loss* yang konsisten hingga *epoch* ketiga, sementara metrik evaluasi tetap stabil di atas 87%, yang menandakan bahwa model memiliki performa klasifikasi sentimen yang baik dengan efisiensi pelatihan yang memadai.



Gambar 11. Hasil Grafik Evaluation Metrics



Gambar 12. Hasil Grafik Training and Validation Loss

Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja *IndoBERT* dalam mengklasifikasikan sentimen pada data uji. Dataset uji terdiri atas 5.023 data atau 20% dari keseluruhan dataset, yang

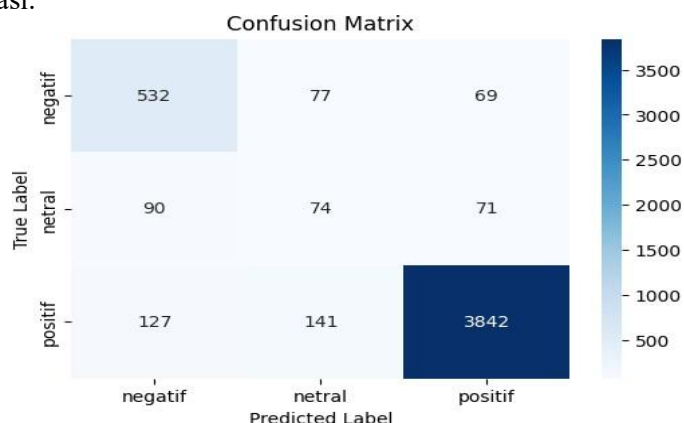
dibagi secara *stratified* sehingga proporsi label positif, netral, dan negatif tetap terjaga. Hasil evaluasi menunjukkan bahwa model mencapai tingkat akurasi sebesar 89%, dengan nilai *F1-score* tertinggi pada kelas positif sebesar 0,95.

Classification Report:

	precision	recall	f1-score	support
negatif	0.71	0.78	0.75	678
netral	0.25	0.31	0.28	235
positif	0.96	0.93	0.95	4110
accuracy			0.89	5023
macro avg	0.64	0.68	0.66	5023
weighted avg	0.90	0.89	0.89	5023

Gambar 13. Hasil Classification Report

Hal ini menunjukkan bahwa model sangat efektif dalam mengklasifikasikan sentimen positif, baik dari sisi *precision* maupun *recall*, yang dipengaruhi oleh dominasi jumlah data positif dalam dataset. Sebaliknya, performa model pada kelas netral relatif rendah dengan nilai *F1-score* sebesar 0,28, yang mengindikasikan adanya kesulitan dalam membedakan sentimen netral dan positif. Kondisi ini merupakan permasalahan umum pada klasifikasi multikelas dengan distribusi data yang tidak seimbang. Untuk menganalisis pola kesalahan prediksi, digunakan *confusion matrix* sebagai alat visualisasi.



Gambar 14. Hasil Confusion Matrix

Berdasarkan hasil *confusion matrix* yang ditampilkan pada Gambar 14, model berhasil mengklasifikasikan 3.842 data positif secara benar, dengan sebagian kecil kesalahan prediksi ke kelas negatif (127 data) dan netral (141 data). Pada kelas netral, model mengklasifikasikan 74 data secara tepat, namun masih terjadi kesalahan prediksi ke kelas negatif (90 data) dan positif (71 data). Sementara itu, pada kelas negatif, sebanyak 532 data berhasil diklasifikasikan dengan benar, dengan kesalahan prediksi masing-masing 77 data ke kelas netral dan 69 data ke kelas positif. Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *IndoBERT* memiliki performa yang sangat baik dalam mengidentifikasi sentimen positif dan negatif, dengan akurasi keseluruhan sebesar 89%. Meskipun demikian, masih terdapat peluang pengembangan lebih lanjut, khususnya dalam meningkatkan performa klasifikasi pada kelas netral, sehingga model dapat digunakan secara lebih optimal dalam analisis sentimen berbahasa Indonesia.

KESIMPULAN

Penelitian ini berhasil mengembangkan visualisasi persebaran penyakit menular di Indonesia periode 2015–2023 melalui penerapan metode *CRISP-DM* yang diintegrasikan dengan algoritma PEDULI (*PEta sebaran penDULu waktu dan Lokasi*). Hasil pengolahan dan visualisasi data dalam bentuk *dashboard* interaktif menggunakan Looker Studio mampu menyajikan informasi pola penyebaran, tren perubahan kasus, serta wilayah dengan tingkat risiko tinggi secara lebih informatif dibandingkan penyajian data konvensional. Penerapan algoritma PEDULI terbukti efektif dalam mendeteksi lonjakan kasus berdasarkan dimensi waktu dan lokasi, sehingga dapat membantu mengidentifikasi penyakit dengan pola peningkatan yang konsisten dari tahun ke tahun. Dengan

demikian, penelitian ini dapat mendukung pemahaman yang lebih komprehensif terhadap dinamika penyakit menular di Indonesia.

Saran

Berdasarkan hasil penelitian, disarankan agar pengembangan sistem visualisasi ini dapat diperluas dengan menambahkan data yang lebih mutakhir serta variabel pendukung lainnya, seperti faktor demografi dan lingkungan, guna meningkatkan ketepatan analisis. Selain itu, integrasi dengan sistem pemantauan kesehatan nasional secara *real-time* diharapkan dapat meningkatkan fungsi *dashboard* sebagai alat pendukung pengambilan kebijakan dan strategi pencegahan penyakit menular di masa mendatang.

DAFTAR PUSTAKA

- Alga, J., Wulandari, C., & Intan, B. (2024). Analisis Sentimen Aplikasi YouTube di Google Play Store menggunakan Machine Learning. *Resolusi*, 4(4). <https://doi.org/10.30865/resolusi.v4i4.1750>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... Wright, R. (2023). "So what if ChatGPT wrote it?": Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Febrywinata. (2024). Pemanfaatan Google Colaboratory sebagai media pembelajaran pemrograman Python berbasis *cloud*. *Jurnal Teknologi Informasi dan Pendidikan*. <https://journal.example.ac.id/google-colab>
- Harun, M., & Ananda, R. (2021). Evaluasi kinerja model klasifikasi menggunakan *confusion matrix*. *Jurnal Teknologi dan Sistem Informasi*, 7(2), 45–52. <https://doi.org/10.31289/jutsi.v7i2.4567>
- Hassani, H., & Silva, E. (2023). The role of ChatGPT in data science and artificial intelligence. *Big Data and Cognitive Computing*, 7(2), 1–12. <https://doi.org/10.3390/bdcc7020062>
- Helmud, E., Fitriyani, F., & Romadiana, P. (2024). Classification comparison performance of supervised machine learning using *confusion matrix* as evaluation. *SISFOKOM*, 13(1). <https://doi.org/10.32736/sisfokom.v13i1.1985>
- Herlinawati, E., Pratama, A., & Nugroho, Y. (2020). Analisis ulasan pengguna aplikasi Android di Google Play Store. *Jurnal Sistem Informasi*, 16(1), 1–10. <https://doi.org/10.21609/jsi.v16i1.950>
- Kalla, D., & Kuraku, S. (2023). ChatGPT-based systems: Architecture, training models, and applications. *Journal of Artificial Intelligence Research and Applications*, 4(1), 21–30. <https://doi.org/10.5430/jair.v4n1p21>
- Marquis, R., Oladoyinbo, A., Olabanji, O., Olaniyi, O., & Ajayi, O. (2024). Transformer-based language models for conversational agents. *International Journal of Computer Science and Information Security*, 22(1), 55–63. <https://doi.org/10.5281/zenodo.10543210>
- Putra, R. S., & Ratih, I. D. (2021). Klasifikasi Tanggapan Pelaksanaan Program Magang dengan Menggunakan Metode Naive Bayes Classifier. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 129–137. <https://doi.org/10.57152/malcom.v1i2.113>

-
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python. *BINA INSANI ICT JOURNAL*, 156–165. <https://doi.org/10.51211/biiict.v7i2.1422>
- Widaad, N. (2024). *Analisis sentimen ulasan pengguna aplikasi ChatGPT di Google Play Store menggunakan metode SVM dan CNN* (Tesis Magister, Universitas Gunadarma). Program Studi Sistem Informasi Bisnis, Magister Manajemen Sistem Informasi.